Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Biomedical Informatics 36 (2003) 177–188

Journal of Biomedical Informatics

www.elsevier.com/locate/yjbin

Representative threats for research in public health surveillance

Michael M. Wagner,[a,*] Virginia Dato,[b,1] John N. Dowling,[a] and Michael Allswede[c]

[a] *The Real-Time Outbreak and Disease Surveillance (RODS) Laboratory, Center for Biomedical Informatics, University of Pittsburgh, Suite 550, 100 Technology Drive, Pittsburgh, PA 15219, USA*

[b] *The Center for Public Health Practice, Graduate School of Public Health, University of Pittsburgh, Forbes-Allies Center, 3901 Forbes Avenue, Suite 210, Pittsburgh, PA 15260, USA*

[c] *Center for Emergency Medicine, University of Pittsburgh, 230 McKee Place, Suite 500, Pittsburgh, PA 15213, USA*

Received 11 March 2003

Abstract

A large number of biological agents can cause natural or bioterroristic disease outbreaks and each can present in a bewildering number of ways (e.g., a few cases versus many cases, confined to a building versus widely disseminated). This 'problem space' is a challenge for designers of early warning systems for disease outbreaks and the sheer size of this space is a barrier to progress.

This paper addresses this problem by deriving nine categories of threats that represent a parsimonious characterization of the problem space. A literature search also identified one or more example outbreaks for each of the nine categories. These outbreaks have occurred in recent times and could be used by researchers in need of actual outbreak data for investigations of the role of different types of surveillance data and algorithms in outbreak detection.

The methodological contribution of this research is a *Criterion Set* of threats for analysis and evaluation of detection systems. This set characterizes the problem space in a tractable manner with less loss of generality than analyses based on one or two selected diseases, which is representative of current analyses.

1. Introduction

Because there are a large number of biological agents that can cause natural or bioterroristic disease outbreaks—each having the ability to present in many different ways (e.g., few cases, many cases, abrupt onset, and gradual onset)—the practice of outbreak detection and investigation are extremely challenging cognitive activities. Designers of information systems that support these activities must manage this complexity during the processes of requirements analysis and specification. For example, designers need to understand for every possible outbreak presentation of each agent how quickly the outbreak must be detected and which kinds of surveillance data and analytics are needed to achieve the requisite timeliness and specificity of detection. A biological agent like anthrax, for example, may cause outbreaks that range in size from one to many cases; are confined to a single building or affect a widely dispersed area; range from classic inhalational presentations to predominantly cutaneous or gastrointestinal presentations; exhibit tempos that range from a sudden, nearly simultaneous appearance of many cases to more protracted outbreaks due to persistence of the source; and involve only humans, only animals, or both.

The sheer size of this 'problem space' is a challenge to developers and researchers and a barrier to progress. Designers must worry that the requirements they use adequately cover the space. As just discussed, the space is multidimensional and includes as axes number of cases, spatial distribution, temporal distribution, vector, organism, presentation, and even environmental conditions. Researchers face similar problems when deciding which threats to study and then in generalizing from their results to the broader questions of detection of important threats.

* Corresponding author. Fax: 1-412-383-8135.
E-mail address: mmw@cbmi.pitt.edu (M.M. Wagner).
[1] Present address: Pennsylvania Department of Health, Division of Infectious Disease Epidemiology, Southwest District Office, Room 514, State Office Bldg., 300 Liberty Avenue, Pittsburgh, PA 15111-1210, USA

1532-0464/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S1532-0464(03)00065-0

There are two possible approaches to design and analysis when faced with this complexity. The first—which is the current practice—is to ignore the complexity by restricting attention to a few threats chosen based on criteria such as urgency or pragmatic research reasons such as availability of data. "Threat lists" promulgated by organizations such as CDC, NATO, and others are one basis for such decisions [1–4]. A key problem with this approach for system designers is that the designer has specifications for a special purpose system only. A problem for the researcher is that the information developed by the study illuminates only a small area of the problem space, oftentimes the same small area. As a result, information available to designers is concentrated in a few small regions around the diseases Influenza [5–11] and Cryptosporidium [12–17] with less information available about other diseases (e.g., Anthrax) [18,19]. There is also illumination around respiratory syndrome [20–22] and gastrointestinal syndrome [22–24] but not around other important syndromes such as *botulism-like*. Large areas of the problem space remain unexplored. There is also little understanding of how results about detection of the studied outbreaks (e.g., Influenza) generalize to detection of other types of outbreaks. A fundamental problem is that many types of outbreaks have not occurred in recent times (or ever) so cannot be studied directly and must be understood by simulation modeling or by analysis of data from similar (surrogate) outbreaks. There has been no published discussion of which threats are similar and therefore can be used as surrogates in research.

The second approach is the one devised for use in this study. This approach attempts to reduce the complexity of the problem space by finding clusters of similar threats. Because our focus is the design of information systems, the clustering of threats is based on similarities in the functional requirements that each threat imposes on a detection system—threats with similar functional requirements for detection are grouped together. The goal of this approach is to develop a parsimonious and tractable characterization of the problem space for the purpose of surveillance system design, one that will facilitate both design of systems and the associated research in public health surveillance that supports system design. The consequence is that a designer feels confident that as long as clear specifications for detection of one threat in a category are met, the resulting system can detect any threat from the category. The potential advantages of a parsimonious characterization are: (1) focusing more attention on a smaller, but more comprehensive set of threats, (2) identifying equivalent threats that may be more amenable to study because of availability of data, and (3) improving tractability of analysis without, ideally, loss of generality.

We created such a set of categories by analyzing the detection requirements of a large set of biological agents identified from several published threat lists. Our analysis clustered the threats into a final set of nine categories, each of which represents a fundamentally different problem in the design of detection systems. We also searched the literature for actual outbreaks of diseases in each category, finding examples from each category that researchers could use in analyses or validation of detection algorithms. We note that our categories—being detection problems such as *building contamination*—differs from existing lists (which are lists of organisms or clinical syndromes such as *anthrax*, *inhalational*).

2. Methods

We started with a large set of biological agents derived from published lists of threats, expanded the set to include consideration of various outbreak presentations of the agents (e.g., number of cases, temporal and spatial variations, disease presentation, and route of transmission), and then clustered the outbreaks and searched the literature for examples from each category.

2.1. Set of biological agents

We created a set of biological agents by merging lists of threats developed by the Centers for Disease Control and Prevention (CDC), the North Atlantic Treaty Organization (NATO), the Defense Threat Reduction Agency (DTRA), and other entities [1–4] (Table 1). These entities created these lists for different purposes ranging from defining reportable conditions, to prioritization of antibiotic stockpiling, to response strategizing.

In the CDC Critical Agent List [4], Category A agents are defined as organisms that pose a risk to national security because they can be easily disseminated or transmitted person-to-person; cause high mortality, with potential for major public health impact; might cause public panic and social disruption; and require special action for public health preparedness. Category B agents are second priority agents and include those that are moderately easy to disseminate; cause moderate morbidity and low mortality; and require specific enhancements of CDC's diagnostic capacity and enhanced disease surveillance. A subset of List B agents includes pathogens that are food borne or waterborne. Category C include emerging pathogens that could be engineered for mass dissemination in the future because of availability; ease of production and dissemination; and potential for high morbidity and mortality and major health impact.

The Defense Threat Reduction Agency and NATO have also developed lists of threats (Table 1) [1,25]. The NATO handbook dealing with potential biological warfare agents lists 31 infectious agents. These are the

Table 1
Merged lists from CDC, DTRA, National Notifiable Diseases, NATO, interview with Russian experts, and USAMRIID

Threat	DTRA bioterrorism list	CDC	NATO	Russian experts top 11 threats	USAMRIID	Reportable list
Anthrax, inhalational	X	A-list	X	Top 4	X	X
Botulism	X	A-list	X	Top 4	X	X
Plague (pneumonic)	X	A-list	X	Top 4	X	X
Smallpox	X	A-list	X	Top 4	X	X
Tularemia (inhaled)	X	A-list	X	X	X	X
Hemorrhagic fever viruses (Omsk, Korean, Ebola, Crimean-Congo, Marburg, Junin, etc).	X	A-list	X	X	X	
Brucellosis	X	B-list	X		X	X
Glanders (Melioidosis)	X	B-list	X	X	X	
Q fever (*C. burnetii*)	X	B-list	X	X	X	
Cholera	X	B-list	X		X	X
Clostridium perfringens (epsilon toxin)	X	B-list	X		X	
Ricin toxin	X	B-list	X		X	
Lassa fever		A-list	X		X	
Yellow fever		C-list	X		X	X
Shigellosis	X	B-list				X
Staphylococcal enterotoxins (B, etc.)	X	B-list	X			
Encephalitis (Russian spring summer, eastern equine, Saint Louis, West Nile, Venezuelan, etc.)	X		X		X	X
Alphaviruses (Venezuelan, eastern, western, equine encephalomyelitis)		B-list			X	
Cryptosporidiosis		B-list				X
Escherichia coli 0157:H7		B-list				X
Salmonellosis		B-list				X
Mycotoxins (trichothecene)	X		X		X	
Rickettsial diseases			X		X	X
Typhoid fever	X		X			X
Venezuelan equine encephalitis			X	X	X	
Hantaviral diseases		C-list			X	
Tick-borne hemorrhagic fever viruses		C-list			X	
Tick-borne encephalitis viruses		C-list			X	
Tuberculosis		C-list				X
Chikungunya fever			X		X	
Diphtheria					X	X
Encephalomyelitis viruses	X				X	
Histoplasmosis			X			X
Influenza				X		X
Marine toxins	X				X	
Palytoxin			X		X	
Psittacosis			X			X
Rocky mountain spotted fever			X			X
Saxitoxin			X		X	
Tetrodotoxin			X		X	
Typhus (epidemic rickettsial)			X	X		
Typhus (scrub)			X	X		
Viral infections					X	X
Western equine			X		X	
Yersinia					X	X

Viral diseases were often grouped using incompatible groupings and we have not attempted to merge those groups. Diseases appearing on only one list: Nipah virus (CDC C-list); Coccidioidomycosis, Dengue (NATO); Machupo (USAMRIID); acquired immunodeficiency syndrome, Amebiasis, *Campylobacter*, Carbon monoxide poisoning, *Chlamydia trachomatis*, Congenital rubella syndrome, food poisoning, Giardiasis, *Hemophilus influenzae* type B (HIB), Hepatitis A, Hepatitis B, Hepatitis C, Kawasaki syndrome, Legionnaires' disease, Leptospirosis, Lyme disease, Lymphogranuloma venereum, Malaria, Measles, Meningitis, Mumps, gonorrhoeae, *Neisseria meningitidis* in blood or csf, Pertussis, Poliomyelitis, Rabies, Reye syndrome, Rheumatic fever, Rubella, Syphilis, Tetanus, Toxic shock syndrome, Toxoplasmosis, Trichinosis (reportable diseases).

representative agents thought to be of military significance by the United States and Western Europe.

A Russian panel of bioweapons experts reviewed microbial agents and concluded that 11 were "very likely to be used" [2]. The four most likely to be used were *Variola*, *Yersinia pestis*, *Bacillus anthracis*, and *Clostridium botulinum*. Also included were *Francisella tularensis*, *Pseudomonas mallei/pseudomallei*, *Rickettsia* spp.,

Coxiella burnetti, Venezuelan equine encephalitis, Marburg, and Influenza viruses.

Notifiable diseases are those diseases that local (state or county) health departments require providers and health care institutions to report. The notifiable diseases included in Table 1 are National Notifiable Infectious Diseases from 2001 [3].

There are other lists of threats and diseases of interest to public health that we did not review. Some are classified such as the DIA summary used by USAMRIID. Others are outside the scope of this report because they are not of central concern to designers of early warning systems for biodefense. These include nosocomial infections, toxicological threats (because the effect is usually obvious immediately and therefore early warning is not a requirement), and chronic diseases.

We merged the lists and sorted the merged list to bring the threats cited by multiple groups to the top using a simple point system that assigned one point per list appearance, an extra point if the threat was on the CDC's B-list, or among the Soviet top 4, and two extra points if it was on the CDC's A-list.

2.2. Clustering procedure

Once we had created the merged and sorted list, we clustered elements from the list as follows: For the first etiologic agent in the merged list, which happened to be Anthrax, we created a category called, temporarily, *Category I* and placed Anthrax in the category (we later renamed the categories based on the commonalities amongst the threats that we had placed within). We then reviewed the types of anthrax outbreaks including building contamination, outdoors, single case, large numbers of cases and others using our knowledge of medicine, infectious disease, and public health (Author V.D. is a public health physician with pediatric infectious disease training, J.D. is a specialist in infectious disease and informatics, M.A. is a board certified toxicologist with military medical training in the area of weapons of mass destruction, and author M.M.W. is a specialist in medicine and informatics) and information in the *Control of Communicable Disease Manual* [26]. We created a detection system diagram showing the types of data needed for early detection of this threat and identified the types (possibly new) of data analytic processing required for early detection.

For each remaining disease in the merged list, we asked whether additional data or analytic components should be added to our diagram of a detection system. For example, if the threat involved the water supply and there was nothing in the diagram referring to analysis of water supply data, then we added water supply data and analysis to the diagram and created a new category for that threat. If no new data or analytics were required, then we added the threat to an existing category based on similarity, from a detection perspective, to other threats already in that category. We made notes of any differences with diseases already in that category—for example that the disease had a vector that was not covered by other diseases already in that category. Once we had completed the clustering of threats into categories, we renamed the categories based on the characteristics of the threats within each category.

In judging whether a threat represented a fundamentally new problem for early detection, we relied on our knowledge of what data could contribute to the detection of the threat. We considered many potential types of data without consideration of whether such data were being utilized by any specific system at present (or even had been considered relevant to early detection of that disease), availability, etc. (Table 2). We relied on our collective research experience with data availability and value, which is described in [27] for the problem of early detection of outbreaks in these determinations.

2.3. Representative threats

After we created the categories, we searched the literature for known outbreaks, bioterrorist events, and

Table 2
Types of data considered in the analysis

(1) Prodromal phase
(a) Over-the-counter sales of pharmaceuticals
(b) Sentinel population monitoring
(c) Social behavior patterns and changes
(d) Social aggregations: families, co-workers, recreational groups
(e) Location of recent travel or gatherings
(f) Food and water intake
(g) Absenteeism
(2) Clinical phase
(a) HMO usage
(b) Outpatient clinical volumes and diagnoses
(c) Medical queries, web-based, Public Health notifications, etc
(d) Medical resource utilization
(e) Emergency Department volume and diagnoses
(f) Culture results-precise microbiological identification
(g) Resistance pattern identification
(h) Radiological reports
(i) Use of non-traditional medicines
(3) Morbidity and mortality
(a) Unusual clinical syndrome occurrence
(b) Unexplained deaths
(c) Poison Center usage
(d) Coroner reports
(4) Allied professional sources
(a) Veterinary illness
(b) Vector surveillance
(c) Industrial theft
(d) Food and water source distribution patterns
(e) Civil engineering data
(f) Weather

outbreaks of emerging infections (listed in a Council of State and Territorial Epidemiologists position paper on emerging infections) [28] to identify one threat in the category that could serve, for research and design purposes, as a representative of the category. We supplemented these sources with a review of infectious disease outbreaks reported in the US in volumes 48 (1999) and 49 (2000) of the Morbidity and Mortality Weekly Report.

As a further analysis, we compared each threat in each cluster with its representative threat, asking whether the threat differed from the representative threat in some way that would bias any design or other analysis (such as a general assessment of national biodefense capacity). For example, if a threat required earlier detection than the representative threat because the underlying disease had a more rapid time course, we noted that an analysis would require not only a determination of the capacity for the representative disease, but also an analytical adjustment for the worst-case disease in the category. In the vector-borne class, for example, an analytical adjustment would be needed for every vector not covered by the vectors for West Nile Encephalitis (the representative disease). One can understand these analytical adjustments within the categories as important variations in the clusters along one or two dimensions.

3. Results

The analysis resulted in nine categories, which represents a dramatic reduction of the problem space. Table 3 lists the nine categories with a brief description of their unique detection requirements, the representative threat, and agents we assigned to the categories. The next sections describe the categories in detail.

3.1. Large-scale bioaerosol

The first category is large-scale bioaerosol release of non-contagious agents. Specifically this category includes intentional, accidental or natural releases of biologic organisms or their toxins capable of affecting many individuals.

Although the diseases in this category present with different clinical syndromes, the data sources required for early detection are similar and the diseases grouped together on this basis. For early detection of threats in this category, a detection system would ideally include components to obtain and process data that are preclinical (e.g., grocery sales) and early clinical (e.g., symptoms and radiological reports). Such syndromic information must be collected and analyzed with attention to corroborating and discriminating data from other sources, such as wind patterns and physical location of individuals in the days preceding onset of illness.

We did not find modern outbreaks of diseases in this category. We selected aerosolized *B. anthracis* to represent this category because of the threat posed by inhalational anthrax, and the existing body of literature and analysis about inhalational anthrax. Respecting the limitations of the conflicting data about the numbers of cases from the 1979 Sverdlovsk release of *B. anthracis* (Kirov strain) from Soviet Biological Weapons Compound 19, there are some broad conclusions that can be made. First, rapid detection of diseases in this category must be given priority because of their lethality. Second, conventionally trained physicians cannot be relied upon to detect rapidly outbreaks in this category. For example, the earliest suspicion of an anthrax outbreak in Sverdlovsk came from autopsy findings (cardinal's cap, a pathognomonic finding in anthrax) [29].

Many biowarfare agents fit this category, because the preferred route of deployment is aerosol dissemination. Some key difference between the agents in this category would be the downwind reach or "footprint" size, infectious dose, lethal dose, incubation period, and duration of prodomal period. In the case of aerosolized *B. anthracis*, a large downwind reach provides the worst-case example.

3.2. Premonitory release of agent

The second category is premonitory, small-scale release of an agent. Specifically, this category includes intentional, accidental or natural infection of one or a limited number of individuals with an unusual organism.

The detection problem here is one of sensitivity for single cases and small outbreaks, not extreme timeliness. To detect a single case, extreme specificity is required and such specificity requires either that the prior probability be extremely high (due to intelligence information) or that the patient's disease has sufficiently declared itself that a physician or pathologist with or without the assistance of specific laboratory tests can confirm the diagnosis. A detection system would require components that scrutinize different sources of data to establish whether individual patients fit fairly specific disease definitions such as the combination of gram positive rods in the blood or cerebrospinal fluid and pneumonia on chest radiograph (which would be suggestive of anthrax). Examples of such surveillance system components include clinical information systems with decision support at the point of care, systems to monitor laboratory reporting of microbiology cultures and free-text processing algorithms that scrutinize autopsy reports, newspaper stories, and obituaries for unusual deaths of animals or humans. The demographics of the victims or the discovery of a geographic clustering of victims could help to identify a common cause. In the absence of astute clinical diagnosis, it is

Table 3
Nine categories, detection requirements, and threats

Categories	Unique detection requirements	Representative outbreak	Other threats
Large-Scale bioaerosol	Rapid detection needed to avert catastrophic morbidity and mortality. Data about early disease manifestations, spatial distribution, weather. Algorithms that can process spatial, temporal, and weather information	1979 Sverdlovsk release of *B. anthracis* (Kirov strain) from Soviet Biological Weapons Compound 19 [29]	Weaponized anthrax, weaponized staph enterotoxin B, weaponized tularemia, weaponized botulism, weaponized *C. burnetii*: Q fever, weaponized *Pseudomonas mallei*: glanders, weaponized *Clostridia perfringens* toxin, Weaponized *Brucella* sp., weaponized, Ricin aerosol, T2 Mycotoxin aerosol, Histo-Coccidioidomycosis
Small premonitory release or contamination	Ability to detect a single case of an unusual disease. Requires high specificity and sensitivity, not earliness. Clinical data and expert physicians	Laboratory-acquired human glanders—MARYLAND, May 2000 [30]	Weaponized anthrax, weaponized staph enterotoxin B, weaponized tularemia, weaponized botulism, weaponized *C. burnetii*: Q fever, weaponized *Pseudomonas mallei*: glanders, weaponized Clostridium perfringens
Enclosure (building, ship) contamination	Knowledge and understanding of effect of building/ship components on distribution of agent	Epidemiologic investigations of bioterrorism-related anthrax, New Jersey, 2001 [32]	Any bioaerosol, Marine Toxin: Saxitoxin, Ciguatoxin, Tetrodotoxin, Palytoxin
Continuous or intermittent release of agent over time	Ability to detect an increased number of cases when viewed over a long time period	*Legionella* Pennsylvania [37,38]	*Legionella pneumophila*, Histo-Coccidioidomycosis, Psittaci: Psittacosis
Contagious person-to-person	Requires data not only about disease, but about social contacts. Modern transportation can result in widely dispersed cases	Outbreak of Influenza A infection among travelers—Alaska and the Yukon Territory, May–June 1999 [61]	Influenza,Variola, Rubeola, Rubella, *Mycobacterium tuberculosis*, Mumps, Smallpox, Diphtheria, *Hemophilus influenzae*, *Mycobacterium leprae*, *meningitidis*, Group A Strep: Rheumatic fever, toxic shock, necrotizing fascitis
Commercially distributed products	Tracking of the food distribution chain and/or social interactions that revolve around food. Rapid detection needed to avert potentially catastrophic morbidity and mortality	A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars [50]	Salmonellosis, *Shigella* sp., *E. coli* 0157, *Brucella* sp., Staph enterortoxin *B. vibrio*, Cholerae, *B. anthracis*, toxic alimentary Aleukia: T2 Mycotoxin, *Clostridium botulinum*: Botulism, Hepatitis A and C. Perfringins E Toxin, Ricin Toxin, "Heavy Metals: Pb, Hg, As," Nipah Virus," Marine Toxin: Saxitoxin, Ciguatoxin, Tetrodotoxin, Palytoxin," *Trichinella*: Trichinosis, Norwalk, Cyclosporiasis
Waterborne	Knowledge and understanding of water distribution systems, rapid detection needed to avert potentially catastrophic morbidity and mortality	Large community outbreak of cryptosporidiosis due to contamination of a filtered public water supply [53]	Cryptosporidiosis, *Shigella* sp., *Camphylobacter*, Giardiasis, Staph enterotoxin B, *E. coli* 0157: "Botulism, bioterroristic," Ricin Toxin, *Entamoeba histolytica*, cyclosporiasis
Vector/host-borne	Transmission vehicles are animals or insects resulting in unusual patterns of cases that cannot be modeled without data and analytics related to hosts and vectors	Outbreak of West Nile-like viral encephalitis—New York [55,56]	Malaria, West Nile, Yellow fever, Dengue, *Yersinia pestis*, Tularemia, Ebola, Marburg, Hantaviruses, Q fever, Glanders, Melioidosis, Lassa, Machupo, Junin, Rift Valley, CCHF, Hantaan, *Alphaviridae*: VEE, EEE, WEE, Chikungunya, *Flaviviridea encephalitis*: Russian spring summer, eastern equine, St. Louis, West Nile, Venezuelan, etc., Nipah Virus, Rabies, Lyme, *Rickettsia* sp.: Rocky Mountain Spotted fever, Typhus
Sexual or parenteral transmission	Knowledge of relationships between cases requires access to highly personal information, tracking of blood supply	Cluster of HIV-positive young women—New York, 1997–1998 [62]	HIV, gonorrhoeae, *Hemophilus ducreyi*, *Treponema pallidum*, *Chlamydia trachomatis*, Hepatitis B and C

likely that a single case of disease caused by a weaponized organism will progress to fatality. The detection problem includes therefore analysis of unexplained deaths.

The representative threat that we identified for this category is human glanders, which is caused by the agent *Burkholderia mallei*. We found a case report of human glanders in Maryland 2000 [30]. This illness—caused by a potential bioterrorism agent (it is on five of the six lists we used to create the merged list)—went undiagnosed for almost two months despite the fact that the victim had an occupational history of working with the agent. Besides glanders, the first cases of inhalational anthrax in October 2001 are also examples from this category worthy of study by system designers.

This type of threat includes almost all organisms and information systems will require specific, computer-interpretable case definitions for each disease of interest. The problem of detecting a single case is identical to the problem of accurate diagnosis in medicine, and there is a large literature on clinical decision support describing relevant techniques [31].

3.3. Building/vessel contamination

The third category is the contamination of a building or other enclosed space with an infectious agent or a biologic toxin. Specifically, this category refers to the distribution of an agent via the mechanical components in a building or ship. This is an important category because of modern reliance upon heating, ventilation, and air-conditioning systems, where victims are linked to a common air source.

The recognition of a release contained within a structure presents a different problem for detection than a large-scale aerosol or premonitory release. Specifically, the detection of a syndrome common to a relatively small number of individuals sharing a domicile, a place of employment, or a social facility requires recognition and analysis of these relationships. A detection system ideally would have access to data about heating, ventilation, and air conditioning systems (HVAC), identity of occupants, and hours of occupation. These requirements are not represented in the previous categories.

For this category, the building contamination with *B. anthracis* of a postal facility in New Jersey is an interesting example for study and validation of detection systems [32]. Tracing the illnesses to the specific building required the knowledge of work times, responsibilities and routines for a large number of postal workers. Although not a biological agent and not on our merged list, carbon monoxide is also a potentially useful representative for this category because of the abundance of data on exposures [33–36]. However, in comparison to other threats in this category, the onset of symptoms of CO poisoning is immediate so that victims typically are found in situ whereas with other threats an incubation period allows for dispersal of cases prior to onset of symptoms. Therefore, if CO poisoning is used analytically to assess requirements for detection of threats in this category, analytical adjustments such as distortion of actual data to achieve artificial post-exposure dispersion of cases prior to symptom onset should be used.

3.4. Continuous or intermittent release of an agent

The fourth category is continuous or intermittent release of an infectious agent or biologic toxin. Specifically, threats in this category involve low level or low virulence continuous or intermittent release of an infectious agent or biologic toxin from a point source so that individuals entering a limited area develop disease over an extended period of time (a large-scale continuous release would be detectable by methods used for category 1—Large-scale Bioaerosol). The release could be intentional, accidental or natural.

Of interest, the Aum Shinrikyo cult used this technique to disseminate continuously liquid anthrax slurry over a period of four days from the rooftop of their building in Tokyo in an attempt to sicken the population of the city [29]. Pets and birds were sickened but no human outbreaks were reported.

Early detection of this type of dissemination would require analysis of syndromic or disease data not only on an instantaneous basis, but also a longitudinal basis; that is, with respect to similar syndromes occurring at abnormally high levels when analyzed over time. This category thus presents a requirement for a different kind of manipulation of epidemiological data such as the use of detection algorithms based on cut point statistics, explicit models of the expected outbreak, or longer analytic periods.

Legionella pneumophila is a representative disease for this category with outbreaks of this type occurring in modern times such as at the Philadelphia Legionnaires Convention [37,38]. This organism was dispersed continuously both prior to and during the convention but its effects on the population went undetected until a sufficiently large number of highly susceptible hosts were exposed. Of note, some recreational water outbreaks would also be informative about this category, since individuals are exposed on an ongoing basis from the same source over time, creating the same temporal pattern of illnesses [39–41].

3.5. Contagious person-to-person

The fifth category is contagious person-to-person transmission of an infectious agent. In particular, this category includes those threats that are spread through causal contact with a person with the disease. Each

individual infected then becomes a potential source for multiple other individuals whose only relationship may be sitting or standing near each other in a public setting.

The recognition of a contagious disease involves detection components described in the previous categories, but also adds a requirement for components capable of collecting and analyzing social network and contact information. Following the initial release or exposure of the index victims, secondary, tertiary, and quaternary victim pools are created for each index case until quarantine is imposed or the entire population is exposed. These successive waves of victims present with similar syndromes, but unlike a continuous point release of a non-contagious agent in the fourth category, the spread of communicable disease follows patterns of social association. In the case of highly contagious airborne diseases, this progression can occur rapidly in populations with close living or working relationships. The identification of a communicable disease and the populations at risk is critical to management as attenuation of social contact or quarantine is of paramount importance.

We identified *Influenza* as the representative disease in this category because it is already the focus of research in early detection cited earlier. There are several other diseases in this group (see Table 3) that occur regularly in outbreak form such as measles [42–46], rubella [42,47,48], and more recently SARS [49].

Analysis of the other diseases in this category revealed that, in practice, Influenza detection is somewhat atypical in that it utilizes patterns of disease on other continents and remote communities that presage local disease (and therefore have implications for control through vaccination). Analytical adjustments of results about general detection capacity based on influenza data might involve blinding the detection system to such data for diseases that do not have this characteristic.

3.6. Commercially distributed products

The sixth category is contamination of commercially distributed products, especially food. Food contamination may be as simple as contamination at the site of preparation, which is what the Rajneesh cult attempted in The Dalles, Oregon, or as involved as tampering with distribution or production facilities [50].

Detection of threats in this category requires analysis of production and distribution systems in addition to the longitudinal analyses required for detection of intermittent and continuous point source releases. Our review of outbreaks in this category also identified the requirement of incorporating data from routine molecular fingerprinting of organisms, which has proven value in the detection of smaller common-cause outbreaks [51,52]. In the simplest case, detection of a food borne pattern requires at least two victims, knowledge of social interactions (to determine a common meal), and analysis of their food consumption. Analysis of food borne outbreaks in settings such as banquets where the victims know each other or the food is consumed in the same location is a routine epidemiological investigation, performed competently in the United States. However, when food is consumed at separate locations and the individuals do not know each other, detection is more difficult. An ideal detection system for diseases in this category would integrate a high degree of behavioral analysis, with sophisticated laboratory analysis and knowledge of food and distribution patterns.

Salmonella sp. is a representative agent in this category. There are many other diseases listed in Table 3 that occur regularly and that could be used to represent the group.

3.7. Waterborne

The seventh category of threats is contamination of a potable water supply (well or surface water). Although American cities enjoy closed source water supplies that minimize simple contamination, large-scale outbreaks due to contamination do occur and these water supplies are still vulnerable to sabotage or deliberate tampering at water treatment and storage facilities. Population centers rely on a relatively small number of water supply facilities; therefore it is theoretically possible (although difficult due to dilutional effects) to expose a large number of individuals with a single act of contamination.

This category presents the requirement that the detection system have components that analyze data about the water supply to assess whether waterborne transmission of a biological agent can explain patterns of illness occurring in the community.

Cryptosporidiosis is the representative threat because of the existence of large outbreaks for study such as ones in Milwaukee and other regions that resulted from contamination of a water supply [12–17,53].

3.8. Vector/host borne

The eighth category is vector and host borne disease. It includes all outbreaks involving the transmission of an agent via an insect vector or animal host. The characteristic epidemic pattern of natural occurrences of diseases in this category is host disease followed by vector disease followed by human disease.

Detection of outbreaks in this category presents a new requirement that the detection system monitor vectors and hosts of the biological agents. For example, detection systems would have to incorporate data from veterinary sources to monitor animal vectors and hosts.

West Nile encephalitis is the representative threat for this category because of its recent occurrence and spread [54–57].

Analysis of the other diseases in this category reveals marked heterogeneity in vectors. Therefore, any analysis of general capacity would have to use an adjunct checklist of vectors (rat, flea, tick, horse, swine, bat, primates, and bird). The existence of an early warning capability for Rift Valley fever based on satellite image analysis suggests that the ability to process environmental data represent an additional requirement for detection systems [58].

3.9. Sexually transmitted

The ninth category of threats is biological agents transmitted through sexual or other intimate contact such as the sharing of needles.

For detection of outbreaks in this category, syndrome recognition, and pathogen identification must be combined with analysis of social and sexual behavior patterns. We differentiated this category from other forms of social and behavioral analysis such as those that may contribute to early detection of epidemics of contagious diseases because the data collection enabling the analysis of sexual behavior patterns is difficult, requires sensitivity, and may infringe on legal rights. A key detection problem raised by this category is identifying a carrier who is infecting other individuals (either intentionally or unintentionally).

We have chosen *human immunodeficiency virus* (HIV) as the representative disease in this category because it best exemplifies the challenges expected to be encountered in this group.

4. Discussion

Progress in science is facilitated when complexity can be reduced without loss of generality. The method described in this paper reduced a high dimensional, very large problem space to nine categories of biological threats. The new typology is not a simple reduction along any single axis—several of the new categories differ along multiple axes including size of outbreak (large cohort versus single case), spatial distribution (in building versus open air), route of dissemination (person-to-person, air, food, water, and vector), and time (sudden versus continuous).

Our approach differs fundamentally from the *list-of-threats* approach in that it was designed specifically to reduce the problem space without loss of generality. Lists used by CDC, NATO, and others merely enumerate threats whereas our categories represent types of threats (e.g., *building contamination*) each of which presents unique requirements for the design of information systems used to detect outbreaks due to the various threats.

The principal contribution of this paper is a methodological one: a suggested criterion set of nine types of threats that designers of surveillance systems can use to develop requirements and specifications. Additionally, we suggest specific examples of disease outbreaks to study and provide references to those outbreaks. We note that several categories have received little attention from researchers concerned with early detection of outbreaks and deserve heightened awareness so that complete requirements for detection systems become available. Such analyses are vitally important to understand the nation's current ability to detect bioterrorism, and to inform the design of early warning systems for disease outbreaks. We recommend that researchers studying these questions consider using the categories we identified, rather than or at least in addition to other lists. We have ourselves used this set in an analysis of the nation's current capacity to detect bioterrorism [59].

Some of our categories—identified by a first principle analysis of the data and data analytic problems posed by the threats—happen to correspond to organizational divisions within public health surveillance (e.g., sexually transmitted, communicable, vector borne, water borne, and food borne). This coincidence may reflect the fact that public health has over time grouped diseases and threats based in part on their similarities from the perspective of detection, especially with respect to the types of data collections and analyses required to achieve greater efficiency through sharing similar surveillance and investigation resources (e.g., epidemiological skills, methods, and knowledge of surveillance personnel) needed by all threats in a category. (It is worth noting that with the advent of information technology, these departmental specializations resulted in undesirable siloing of information systems and that the authors of this paper recommend that the specific requirements identified for each of the nine categories be pooled to create an overall set of requirements for an integrated approach to surveillance.) It is also possible that our research team was influenced by an anthropomorphic tendency to assume that automatic systems will use the same data and analytic approaches as humans, although we think this effect is unlikely because we used very different assumptions about the types of data that would be available for analysis. For example, we assumed that many types of data that are not ordinarily collected by surveillance systems (they are only collected during investigations) could be made available routinely for surveillance through new information systems. We assumed that variables typically ascertained by investigators manually could be derived automatically and possibly indirectly by inference from existing data sources.

Each of these nine categories represents an important and different problem in detection. It is difficult to prove a claim that such a radical compression of the problem space can be achieved without any loss of generality, and in fact mail dissemination as a route of dissemination was not identified by our analysis (which was conducted prior to October 2001), so the authors do not wish to make that claim. A defensible claim is that this new set of criterion threats is a more complete characterization than current approaches. One can ask, for example, whether this set will be more useful to designers than alternative sets such as the CDC A-list—comprising inhalational Anthrax, Botulism, Plague (pneumonic), Smallpox, Tularemia (inhaled), Hemorrhagic fever viruses (Omsk, Korean, Ebola, Crimean-Congo, Marburg, Junin), and Lassa fever. From a designer's perspective, the A-list is redundant for aerosol releases (anthrax, tularemia, and plague). It is not explicit about the design requirements related to detection of single cases and building contaminations. Although recent experience with anthrax makes those requirements implicit, they might be overlooked by designers. Similarly, it is not explicit that *botulism* maps into design requirements for the detection of outbreaks caused by both food and water supply contaminations. The A-list does not consider continuous releases and sexually transmitted methods of dissemination and would not encourage a systematic examination of similar threats in each category to ensure that additional design requirements were identified. In general, the threat lists are disproportionately weighted with diseases from the aerosolized/weaponized category suggesting that such threat lists do not represent fully the problem space. This would imply caution in using such threat lists alone for guiding development or assessment of detection capacity.

We note that alternative characterizations of the problem space are possible. For example, Wagner [60] describes a division into two parts. The first category was cohort exposures, in which many individuals are infected nearly simultaneously through contamination of the air, food, or water. These individuals move through the phases of illness together as a cohort and it is essential that they be detected at the early syndromic phase through spatial and temporal analysis of syndromic data if the outbreak is to be detected early enough to mitigate mortality and morbidity. The second category was outbreaks that present initially with few cases such as smallpox or any threat that falls into the above *premonitory release of agent* category. The latter category will be dependent on astute clinicians or pathologists not missing the first case. The two category division of the problem space is useful because it suggests that both syndromic approaches and approaches based on improving the diagnostic and management abilities of front-line clinicians are *both* necessary, but it does not provide sufficient detail for designers of early warning systems.

Important questions for future study include further defining the detection requirements for each category. The rational development of public health surveillance systems depends on more exact specification of the timeliness requirements created by the aggressiveness of specific diseases. A timeliness analysis is best paired with a cost estimation of the costs of false alarms. Such analyses will form a framework for understanding which data are needed to detect each threat and the design problems associated with obtaining such data within that time requirement.

A second important question for future study is an extension of our analysis to toxicological, nosocomial, and possibly chronic diseases because of the dual and triple use of detection systems for these purposes.

5. Conclusions

The very large numbers of disease outbreaks can be clustered into nine categories, which represent fundamentally different problems in outbreak detection. There are recent examples for almost all categories for researchers to study, with the exception of the pattern "large-scale aerosol release." These categories can be used as a Criterion Set against which the nation's capacity to detect both naturally occurring and bioterroristic disease can be analyzed.

Acknowledgments

This work was supported by Contract No. 290-00-0009 from the Agency for Healthcare Research and Quality; and National Library of Medicine Applied Informatics Fellowship, Grant No. 5 F38 LM00089. The authors thank Dr. William Hogan for insightful comments.

References

[1] NATO Handbook on the Medical Aspects of NBC Defensive Operations AMedP-6(B). NATO; 1996.
[2] Henderson DA. The looming threat of bioterrorism. Science 1999;283(5406):1279–82.
[3] Roush S et al. Mandatory reporting of diseases and conditions by health care professionals and laboratories. JAMA 2001;282(2):164–70.
[4] Office of the Assistant Secretary for Planning and Evaluation, O.o.S.P., FY 2002–FY 2006 plan for combating bioterrorism. Washington: Department of Health and Human Services; 2001.
[5] Irvin CB, Nouhan PP, Rice K. Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity for bioterrorism and influenza surveillance. Ann Emerg Med 2003;41(4):447–52.

[6] Mostashari F et al. Use of ambulance dispatch data as an early warning system for communitywide influenzalike illness, New York city. J Urban Health 2003;80(Suppl 1):I43–9.
[7] Tsui F-C et al. Value of ICD-9-coded chief complaints for detection of epidemics. Proc AMIA Symp 2001:711–5.
[8] Quenel P et al. Sensitivity, specificity and predictive values of health service based indicators for the surveillance of influenza A epidemics. Int J Epidemiol 1994;23(4):849–55.
[9] Quenel P, Dab W. Influenza A and B epidemic criteria based on time-series analysis of health services surveillance data. Eur J Epidemiol 1998;14(3):275–85.
[10] Toubiana L, Flahault A. A space–time criterion for early detection of epidemics of influenza-like-illness. Eur J Epidemiol 1998;14(5):465–70.
[11] Le Strat Y, Carrat F. Monitoring epidemiologic surveillance data using hidden Markov models. Stat Med 1999;18(24):3463–78.
[12] Naumova EN et al. The elderly and waterborne Cryptosporidium infection: gastroenteritis hospitalizations before and during the 1993 Milwaukee outbreak. Emerg Infect Dis 2003;9(4):418–25.
[13] Stirling R, et al. North Battleford, Saskatchewan, Spring 2001 waterborne Cryptosporidiosis outbreak, part 2. Health Canada; 2001.
[14] Stirling R et al. Waterborne Cryptosporidiosis outbreak, North Battleford, Saskatchewan, Spring 2001. Can Commun Dis Rep 2001;27(22):185–92.
[15] Mac Kenzie WR et al. A massive outbreak in Milwaukee of Cryptosporidium infection transmitted through the public water supply. N Engl J Med 1994;331(3):161–7.
[16] Corso PS et al. Cost of illness in the 1993 waterborne Cryptosporidium outbreak, Milwaukee, Wisconsin. Emerg Infect Dis 2003;9(4):426–31.
[17] Proctor ME, Blair KA, Davis JP. Surveillance data for waterborne illness detection: an assessment following a massive waterborne outbreak of Cryptosporidium infection. Epidemiol Infect 1998;120(1):43–54.
[18] Goldenberg A et al. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. Proc Natl Acad Sci USA 2002;99(8):5237–40.
[19] Kaufmann A, Meltzer M, Schmid G. The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable? Emerg Infect Dis 1997;3(2):83–94.
[20] Espino J, Wagner M. The accuracy of ICD-9 coded chief complaints for detection of acute respiratory illness. Proc AMIA Symp 2001:164–8.
[21] Lazarus R et al. Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. BMC Public Health 2001;1(1):9.
[22] Lewis M et al. Disease outbreak detection system using syndromic data in the greater, Washington, DC area. Am J Prev Med 2002;23(3):180.
[23] Ivanov O et al. Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance. Proc AMIA Symp 2002:345–9.
[24] Pavlin JA. Investigation of disease outbreaks detected by "Syndromic" surveillance systems. J Urban Health 2003;80(Suppl 1):I107–14.
[25] Department of Defense, Chemical and biological defense program: annual report to congress; 2000.
[26] Chin J, editor. Control of communicable diseases manual. American Public Health Association; 2000.
[27] Wagner MM, Aryel R, Dato V. Availability and comparative value of data elements required for an effective bioterrorism detection system. Washington, DC: Agency for Healthcare Research and Quality; 2001. p. 184.
[28] Osterholm M. CSTE position statement 1993-2; 1993.
[29] Mangold, Goldberg. Plague wars. St. Martin's Press; 1999.
[30] Anonymous, Laboratory-acquired human glanders—Maryland, May 2000. MMWR Morb Mortal Wkly Rep 2000;49(24):532–5.
[31] Miller RA. Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. J Am Med Inform Assoc 1994;1(1):8–27.
[32] Greene CM et al. Epidemiologic investigations of bioterrorism-related anthrax, New Jersey, 2001. Emerg Infect Dis 2002;8(10):1048–55.
[33] Anonymous, Carbon monoxide poisoning associated with use of LPG-powered (propane) forklifts in industrial settings—Iowa, 1998. MMWR Morb Mortal Wkly Rep 1999;48(49):1121–4.
[34] Anonymous, From the Centers for Disease Control and Prevention. Houseboat-associated carbon monoxide poisonings on Lake Powell—Arizona and Utah, 2000. JAMA 2001;285(5):530–2.
[35] Anonymous, Houseboat-associated carbon monoxide poisonings on lake Powell—Arizona and Utah, 2000. MMWR Morb Mortal Wkly Rep 2000;49:1105–8.
[36] Anonymous, From the Centers for Disease Control and Prevention. Carbon monoxide poisoning deaths associated with camping—Georgia, March 1999. JAMA 1999;282(14):1326.
[37] Anonymous, Respiratory infection—Pennsylvania. MMWR 1976:244.
[38] Anonymous, Follow-up on respiratory disease—Pennsylvania. MMWR 1976:267.
[39] Tillett HE, de Louvois J, Wall PG. Surveillance of outbreaks of waterborne infectious disease: categorizing levels of evidence. Epidemiol Infect 1998;120(1):37–42.
[40] Barwick RS et al. Surveillance for waterborne-disease outbreaks–United States, 1997–1998. Morb Mortal Wkly Rep. CDC Surveillance Summaries 2000;49(4):1–21.
[41] Anonymous, Outbreak of gastroenteritis associated with an interactive water fountain at a beachside park—Florida, 1999. MMWR Morb Mortal Wkly Rep 2000;49(25):565–8.
[42] Doyle TJ, Bryan RT. Infectious disease morbidity in the US region bordering Mexico, 1990–1998. J Infect Dis 2000;182(5):1503–10.
[43] Anonymous, Measles outbreak—Netherlands, April 1999–January 2000. MMWR Morb Mortal Wkly Rep 2000;49(14):299–303.
[44] Mary M et al. Six years of public health surveillance of measles in France. Int J Epidemiol 1992;21(1):163–8.
[45] Nobre FF, Stroup DF. A monitoring system to detect changes in public health surveillance data. Int J Epidemiol 1994;23(2):408–18.
[46] Valleron AJ et al. A computer network for the surveillance of communicable diseases: the French experiment. Am J Public Health 1986;76(11):1289–92.
[47] Anonymous, Rubella outbreak—Westchester County, New York, 1997-1998. MMWR Morb Mortal Wkly Rep 1999;48(26):560–3.
[48] Anonymous, Rubella among hispanic adults—Kansas, 1998, and Nebraska, 1999. MMWR Morb Mortal Wkly Rep 2000;49(11):225–8.
[49] Lee N et al. A major outbreak of severe acute respiratory syndrome in Hong Kong. N Engl J Med 2003;348(20):1986–94.
[50] Torok TJ et al. A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. JAMA 1997;278(5):389–95.
[51] PulseNet, The National Molecular Subtyping Network for Foodborne Disease.
[52] Degener JE et al. Nosocomial infection by *Staphylococcus haemolyticus* and typing methods for epidemiological study. J Clin Microbiol 1994;32(9):2260–5.
[53] Hayes E, MT, O'Brien T, McKinley T, Logsdon D, Rose J, et al. Large community outbreak of Cryptosporidiosis due to contamination of a filtered public water supply. N Engl J Med 1989;320(21):1372–6.

[54] National West Nile Virus (WNV) Surveillance System.
[55] Anonymous, Outbreak of West Nile-like viral encephalitis—New York, 1999. MMWR Morb Mortal Wkly Rep 1999;48(38):845–9.
[56] Anonymous, Update: West Nile Virus Encephalitis—New York, 1999. MMWR 1999;48(41):944–6, 955.
[57] Fine A, Layton M. Lessons from the West Nile viral encephalitis outbreak in New York City, 1999: implications for bioterrorism preparedness. Clin Infect Dis 2001;32(2):277–82.
[58] GEIS, GEIS Rift Valley fever monitor; 2001.
[59] Dato V et al. The nation's current capacity for the early detection of public health threats including bioterroris agency for healthcare research and quality. Washington: Agency for Healthcare Research and Quality; 2001. p. 78.
[60] Wagner M, et al. The role of clinical information systems in public health surveillance. In: Ball M, Weaver C, Kiel J, editors. Healthcare information management systems, 3rd ed., New York: Springer; 2004, in press.
[61] Anonymous, Outbreak of Influenza A infection among travelers—Alaska and the Yukon Territory, May–June 1999. MMWR Morb Mortal Wkly Rep 1999;48(25):545–6, 555.
[62] Anonymous, Cluster of HIV-positive young women—New York, 1997–1998. MMWR Morb Mortal Wkly Rep 1999;48(20):413–6.